

Der „Dr. Fox“ – Effekt

Oder: Welchen diagnostischen Wert haben Bewertungen von Vorlesungen?

Uwe Füllgrabe

Magazin für die Polizei, 30. Jahrgang, Nr. 273 - 274, Januar/ Februar 1999, S.19 - 26

Zusammenfassung

Nicht nur durch andere Menschen kann man in die Irre geführt werden. Man kann auch leicht durch Informationen getäuscht werden und beispielsweise von nichtssagenden oder sogar falschen Informationen begeistert sein. Deshalb ist es notwendig, Kriterien dafür zu entwickeln, Informationen kritisch zu bewerten. Wie wichtig dies ist, zeigt exemplarisch der „Dr. Fox – Effekt“.

1. Subjektive Einflüsse bei der Bewertung von Informationen

Um die Qualität von Vorlesungen, Seminaren und anderen Unterrichtsformen und die Qualität des Lehrverhaltens zu bewerten, wird häufig die Bewertung von Schülern und Studenten herangezogen. Stellen diese Einschätzungen aber objektive Bewertungen dar? Sagen sie objektiv etwas über die Qualität einer Schulstunde, Vorlesung oder die Lehrbefähigung aus?

Verschiedene Untersuchungen zeigen, dass dies nicht der Fall ist, dass Ablehnung oder Begeisterung hinsichtlich eines Referenten/Vortrages usw. oft unabhängig von dem Inhalt und seiner Qualität ist. Gründe dafür können sein:

- Sympathie oder Antipathie des Beurteilers für den Vortragenden
- Die Erwartungseinstellung: Der Vortrag oder der Artikel einer Person, die positiv gesehen wird, die sympathisch ist oder als kompetent angekündigt wurde, wird automatisch besser beurteilt als der gleiche Vortrag oder Artikel, von dem man glaubt, er sei von einer negativ bewerteten Person (unsympathisch, als inkompetent usw. angesehen) (s. Füllgrabe, 1978).
- Konzentration auf die äußere Form der Darbietung und Vernachlässigung des Inhalts. Begeisterung ersetzt kritische Informationsaufnahme.

1.1 Die positive Bewertungen inhaltsarmer Vorträge

Naftulin, Ware & Donnelly (1973) gingen von folgender Hypothese aus: Wenn man ein genügend eindrucksvolles Paradigma (wissenschaftliches Modell) bei einer Vorlesung benutzt, wird sich eine erfahrene Gruppe von Ausbildern/Erziehern, die an einer neuen Lernsituation teilnehmen, sich zufrieden fühlen, dass sie etwas gelernt haben, trotz irrelevantem, widersprüchlichem und bedeutungslosem Inhalt, den der Vortragende vermittelte.

Um diese Hypothese zu testen, wählten sich die Autoren einen professionellen Schauspieler, der herausragend auftrat und autoritativ klang. Sie versahen ihn mit einem genügend mehrdeutigen Titel, *Dr. Myron Fox*, eine Autorität der Anwendung von Mathematik auf menschliches Verhalten. Man schmückte ihn mit einem ausgedachten aber eindrucksvollen Lebenslauf und stellte ihn drei Gruppen von Personen aus Bereichen des Erziehens vor.

Das Thema von Dr. Fox war „Mathematische Spieltheorie und ihre Anwendung auf körperliche Erziehung“. Als Quellenmaterial diente ein komplexer, aber genügend verständlicher wissenschaftlicher Artikel, der sich an Laien wandte. Doch "Dr. Fox" benutzte viele mehrdeutige Aussagen, Wortneuschöpfungen, unlogische, widersprüchliche Aussagen. All dies war eingebettet in gelegentlichen Humor und bedeutungslose Hinweise auf beziehungslose Themen.

Gruppe I

11 Psychiater, Psychologen und Sozialarbeiter, die sich zu einer Konferenz über Weiterbildung zusammengefunden hatten. Der Zweck der Weiterbildung war, dieser Gruppe zu helfen, bessere Ausbilder für andere Personen aus Heilberufen zu werden, indem sie ihnen verschiedene Unterrichtsziele, Medien und Erfahrungen vermittelte. Dr. Fox hielt eine Stunde lang eine Vorlesung vor ihnen und dann eine halbstündige Diskussion, die auch nicht inhaltsreicher war.

Die Gruppe gab eine Reihe positiver Bewertungen über Dr. Fox ab (Naftulin et al., 1973, p. 632). „Ausgezeichnete Darstellung, ich hörte ihm gern zu.“ „Hat eine gründliche Art“. „Eine viel zu intellektuelle Darbietung“. „Meine Orientierung ist pragmatischer“.

Obwohl also auch einige kritische Bewertungen vorkamen, überwogen statistisch signifikant und deutlich die positiven Bewertungen.

Gruppe II

Eine zweite Gruppe bestand aus 11 Psychiatern, Psychologen und psychiatrischen Sozialarbeitern. Auch hier überwogen deutlich die positiven Bewertungen. Einer behauptete sogar, die Veröffentlichungen des Vortragenden gelesen zu haben. Aber es gab auch kritische Stimmen, was anzeigt, dass es auch Menschen gibt, die sich nicht durch die Darstellungsweise blenden lassen:

"Führte es nicht weit genug. Mangel am bildlichen Material, um es auf die Psychiatrie zu beziehen. Zuviel Gestik. Ließ bedeutsame Beispiele aus. Er ließ die letzten paar Sätze aus, die meiner Ansicht nach seine Ideen für mich verbunden hatten" (Naftulin et al., 1973, p. 632).

Gruppe III

Die dritte Gruppe bestand aus 33 Erziehern und Verwaltern, die sich in einem Kurs für Erziehungsphilosophie eingeschrieben hatten.

Auch hier überwogen signifikant die positiven Bewertungen. Aufschlussreich sind einige subjektive Bewertungen:

„Lebendige Beispiele. Die entspannte Art der Darbietung war ein starker Faktor, der mein Interesse aufrechterhielt. Extrem guter Ausdruck. Interessant, hätte mir gewünscht, dass er sich mehr auf Hintergrundinformationen bezog. Gibt Analyse des Sachverhalts, den ich persönlich vorher studiert hatte. Sehr dramatische Darbietung. War gewiss fesselnd“.

Aber es gab auch kritische Stimmen: „Etwas desorganisiert. Frustrierend langweilig. Unorganisiert und unwirksam“.

Dies zeigt auf, dass der gleiche Vortrag von verschiedenen Personen höchst unterschiedlich bewertet wird. Offensichtlich kann man also zu den meisten Vorträgen sowohl positive als auch negative Stimmen hören.

Zuhörer eines Vortrages - selbst professionelle Experten im Ausbildungswesen - können wirkungsvoll zu der Illusion verleitet werden, etwas gelernt zu haben, wenn der Vortragende einen Stil von Autorität und Witz simuliert.

Portman und Weingartner hatten darauf hingewiesen, dass das Kriterium für einen guten Entdecker von "Mist" darin besteht, dass er nicht vollkommen von den „willkürlichen Abstraktionen der Gemeinschaft eingefangen ist, in der er aufgewachsen ist.“ Natürlich ist eine ausgewogene Kombination von Wissen und anschaulicher Vermittlung sinnvoll, aber nicht, wenn es nur zur Illusion führt, etwas gelernt zu haben. „... zum Lehren gehört mehr, als Schüler glücklich zu machen“ (Naftulin et. al., 1973, p. 634).

Nachdem die Mitglieder der drei Gruppen über den wahren Zweck der Vorlesung informiert worden waren, äußerten einige Interesse, mehr über das angesprochene Gebiet zu lernen. Dies kann einerseits positiv, als Interesse an einem neuen Thema, gesehen werden. Naftulin et al. (1973, p. 631) äußerte aber auch ihren Eindruck, dass die "willkürlichen Abstraktionen" von Postman und Weingartner eine zündende „Aufpumpwirkung“ auf Lernmotivation haben.

1.2 Begeisterung trotz geringer inhaltlicher Substanz

Verschiedene Untersuchungen zeigen also, dass viele Zuhörer nicht in der Lage sind, den Mangel an inhaltlicher Substanz in Vorlesungen zu erkennen, bei denen der Vortragende sehr ausdrucksstark auftritt. In diesen Fällen bot ein als "Experte" eingestufte Vortragende vage Allgemeinplätze über ein Thema der, das den Zuhörern unvertraut war. Goffman (1959) hatte vorhergesagt, dass das Ausdrucksverhalten eine Zuhörerschaft genauso beeinflusst wie der Inhalt des Vortrags, besonders wenn die Zuhörerschaft wenig Zeit oder gedankliche Verarbeitungsmöglichkeit hat, um die dargebotene Information zu bewerten. Williams und Ware (1976) zeigten, dass dies aber sogar dann nicht der Fall ist, wenn Zuhörer viele Möglichkeiten haben, den Informationsgehalt objektiv einzuschätzen. Unmittelbar nach der Bewertung der 1. Vorlesung (Biochemie des Lernens) besuchten die Studenten eine 2. Vorlesung über "Auswirkungen der Stimulation des Gehirns durch die Umwelt". Auch hier zeigte sich die überwältigende Auswirkung der mit starker Ausdruckskraft vorgetragenen Vorlesungen. Die Studenten hatten Dr. Fox nicht besser durchschaut: Auch Vorlesungen mit geringen Inhalt wurden gut bewertet, wenn sie mit großer Ausdruckskraft vorgetragen wurden!

1.3 Der Einfluss von materiellen Leistungsanreizen

Wie bewerten aber die Studenten den Vortrag, wenn sie etwas aus ihm lernen müssen? Unter diesen Bedingungen wäre zu erwarten, dass sie sensibler für die Information sind und den Vortrag genauer bewerten. Williams und Ware (1976) untersuchten diese Frage mit 6 Vorlesungen auf Video über die Biochemie des Lernens und Gedächtnisses. Diese Videos beinhalten zwei Faktoren in unterschiedlichem Ausmaß:

Informationsgehalt, Inhalt und Grad des Ausdrucks bei der Darbietung. Für das Video mit hohem Informationsgehalt wurden 26 lernrelevante Gesichtspunkte in den Vortrag eingearbeitet. Das Video mit dem mittleren Informationsgehalt enthielt 14 und das mit geringem Informationsgehalt 4 Gesichtspunkte. Damit die Videos trotz des unterschiedlichen Informationsgehaltes gleiche Länge (19,1 - 21,8 Minuten) hatten, wurden die mit mittleren und geringen Informationsgehalt ausgefüllt mit

- a) detaillierten Beschreibungen experimentellen Vorgehens ohne Ergebnisse zu liefern,
- b) Diskussionen von Beispielen, Ereignissen und Geschichten, die keinen Bezug hatten,
- c) zeitraubende Versicherungen, dass man bestimmte Dinge später noch ansprechen wolle, ohne dies zu tun,
- d) zirkuläre Diskussion unwichtiger Begriffe und Äußerungen.

Der bereits in früheren Untersuchungen aufgetretene Schauspieler Michael Fox bot diese drei Vorlesungen mit starker oder geringer Ausdruckskraft dar. Eine hohe "Persönlichkeit", Ausdruckskraft beinhaltete: Enthusiasmus, Humor, Freundlichkeit, Charisma.

Zusätzlich zu dem Verhalten des Schauspielers waren Witze und humorvolle Geschichten in die Drehbücher der Vorlesung mit hoher Ausdruckskraft eingefügt worden.

Studenten, die *vor* der Vorlesung einen Leistungsanreiz erhielten, gaben systematisch geringere Bewertungen zu allen Vorlesungen als Studenten, die nach der Vorlesung den Leistungsreiz erhielten. Doch spiegelten die Bewertungen der Studenten in den Vorlesungen mit hoher Ausdruckskraft, die *vorher* Leistungsreize erhielten, nicht den Informationsgehalt der Vorlesungen wider.

Das bedeutet, dass selbst dann, wenn Lernanreize (Geld, Noten usw.) vorhanden sind, die Studenten nicht sensibler für die Qualität der Vorlesung waren.

Je mehr Informationen vermittelt wurden, desto besser die Leistung der Studenten. Die Ausdrucksfähigkeit hatte keinen Einfluss auf die Leistung. Studenten, die vor der Lektüre einen äußeren Leistungsanreiz erhalten hatten, verbesserten ihre Leistung im Vergleich mit Studenten, die erst nach der Vorlesung Leistungsanreize angeboten bekamen.

Wurde die Vorlesung ohne viel Ausdruckskraft vorgetragen, war die Zufriedenheit der Studenten signifikant geringer bei einem geringen Informationsgehalt. Wurden die Vorlesungen mit viel Ausdruckskraft vorgetragen, ergaben sich hohe Unterschiede der Zufriedenheit für die 3 Informationsgrade. Starke Ausdruckskraft kann also einen geringen Informationsgehalt überdecken. Nur wenn die Vorlesung in einem Stil geringer Ausdruckskraft gehalten wurde, war die Zufriedenheit der Studenten mit der Vorlesung mit geringem Informationsgehalt geringer als bei den mit mittlerem oder hohem Informationsgehalt.

Eine starke Ausdruckskraft des Vortragenden erzeugt auch bei informationsärmeren Vorlesungen Zufriedenheit.

1. 4 Zusammenfassung der "Dr. Fox-Studien"

- 1) Die Einschätzungen durch Studenten spiegeln nicht den Informationsgehalt der Vorlesung wider, obwohl der Informationsgehalt die Leistung direkt beeinflusst.
- 2) Das Sensibilisieren von Studenten für den Inhalt der Vorlesungen durch einen materiellen Lernanreiz für das Lernen aus der Vorlesung verbesserte nicht die Genauigkeit der Einschätzung des Informationsgehaltes der Vorlesung.
- 3) Die Einschätzung des Informationsgehaltes verbesserte sich nicht, wenn die Studenten eine zweite Vorlesung unter gleichen Bedingungen erhielten. Selbst bei einer zweiten Vorlesung zu einem anderen Thema durchschauten viele Studenten den "Dr. Fox-Effekt" nicht.

- 4) Studenten erwiesen sich als gute Beobachter des Grades der Ausdrucksstärke eines Vortragenden.
- 5) Studenten sind keine genauen Beobachter zweier wichtiger Gesichtspunkte der Fähigkeiten von Lehren:
 - a) Unterschiede in der Informationsvermittlung von Vortragenden.
 - b) Unterschiede ihrer eigenen Leistung als Ergebnis des Vortrages.
- 6) Vortragende mit hoher Ausdruckskraft wurden besser von den Studenten bewertet als solche mit geringer Ausdruckskraft.
- 7) Studenten, die eine mit hoher Ausdruckskraft vorgetragene Vorlesung sahen, aber kaum äußeren Anreiz zum Lernen erhielten, erbrachten bessere Leistungen als Studenten, denen eine Vorlesung mit geringer Ausdruckskraft vorgetragen wurde.
- 8) Studenten, die Vorlesungen mit großem Informationsgehalt hörten erbrachten bessere kognitive Leistungen als diejenigen, die Vorlesungen mit geringem Informationsgehalt erlebt hatten.

2. Unverständliche Texte werden besser bewertet!

Aus pädagogischer Sicht ist es sinnvoll, einen Text oder eine Vorlesung verständlich zu gestalten. Doch zeigt sich in der Praxis immer wieder, dass oft schwerer verständliche Darstellungen eine höhere Bewertung erfahren als einfachere, leicht verständliche Texte.

Offensichtlich lassen sich viele Zuhörer/ Leser von der Darstellungsform mehr beeindrucken als vom Inhalt, wie ja auch der „Dr. Fox – Effekt“ zeigte.

Vielleicht wirkt folgender Mechanismus :der Leser eines schwerverständlichen Textes oder der Zuhörer eines entsprechenden Vortrages geht davon aus, dass der Autor/ Vortragende schon genau wisse , was er schreibe/sage ,und nur er, der Leser/Zuhörer sei nicht so klug, nicht auf dem neuesten Wissensstand, oder wegen eines anderen Handicaps nicht in der Lage, den tieferen Sinn der Aussagen zu erfassen. Die Parallele zu „Wahrsagern“ soll hier nicht unerwähnt bleiben. Denn ein Wahrsager wirkt deshalb so überzeugend und ist deshalb erfolgreich, weil er seinem Klienten die Bürde der Deutung zuschiebt. Entwickeln sich die Dinge anders als „geweissagt“, so lag es an der fehlerhaften Deutung des *Klienten* (Hymans, 1977). Zum Problem der Schwerverständlichkeit stellte der Wirtschaftswissenschaftler J.S. Armstrong folgende These als Variante des „Dr. Fox – Phänomens“ auf : „Eine unverständliche Botschaft aus einer nach Meinung des Empfängers glaubwürdigen Quelle wird dazu führen, dass der Autor dieser Aussage für besonders kompetent gehalten wird.“

„Armstrong bat 20 Professoren der Wirtschaftswissenschaft, das akademische Prestige von zehn Fachzeitschriften ihres Bereiches zu bewerten. Diese Zeitschriften waren vorher mit Hilfe eines Lesbarkeits-Tests (Flesch Reading Ease Test) in verschiedene ‚Konsumierbarkeits- Grade‘ eingeteilt worden. Das Ergebnis fiel aus wie erwartet: Die Zeitschrift, der man das höchste Prestige zusprach, war am unverdaulichsten: am wenigsten schätzten die Beurteiler eine Zeitschrift, deren Artikel in leicht verständlicher Form abgefasst waren.

Aber kann es nicht auch so sein, dass die Zeitschriften mit höherem Prestige sich komplizierter Themen annehmen und daher eine schwierigere Sprache erfordern? Armstrong untersuchte diese Möglichkeit, indem er Abschnitte aus Artikeln betriebswirtschaftlicher Fachzeitschriften umschrieb, um sie lesbarer zu machen, ohne den Inhalt zu verändern. So verzichtete er beispielsweise auf unnötige Begriffe,

ersetzte schwierige Ausdrücke durch leichter verständliche, löste lange Sätze in mehrere kurze auf.

Eine Passage lautete beispielsweise im Original: ‚Dieser Beitrag führt zu der Schlussfolgerung, dass, um die Wahrscheinlichkeit zu erhöhen, dass ein (Bank-) Kunde sich in eine Schlange einreihet, der Angestellte versuchen sollte, die anfängliche subjektive Einstellung des Kunden hinsichtlich der durchschnittlichen Bedienungszeit dahingehend zu beeinflussen, ihm den Eindruck zu vermitteln, sie sei kurz, oder den Kunden davon zu überzeugen, dass sein eigener Zeitwert der Bedienung groß ist.‘

Armstrongs Fassung lautete: ‚Sie können eher erreichen, dass ein (Bank-)Kunde in einer Schlange wartet, wenn sie ihm den Eindruck vermitteln, dass er nicht lange warten muss. Eine andere Möglichkeit besteht darin, dass es sich für ihn auszahlt, zu warten.‘

Vier solche Passagen gab Armstrong in jeweils leichter oder komplizierter Version einer weiteren Versuchsgruppe von 32 Wirtschaftswissenschaftlern, die er bat, auf einer von 1 bis 7 reichenden Skala ‚die Kompetenz der Forschung, über die in diesem Artikel berichtet wird‘ einzuschätzen. Die Versuchspersonen erfuhren dabei weder den Namen der Zeitschrift, noch den der Autoren.

Wieder einmal beurteilten die Professoren die leicht verständlichen Texte schlechter als die schwierigen.“ (Psychologie Heute: „Reden Sie mal höheren Blödsinn“, August 1980, S.7 - 8).

3. Die unterschiedliche Bewertung wissenschaftlicher Artikel

3.1 Die geringe Übereinstimmung von Gutachtern

Auch der Prozess des Bewertens literarischer und wissenschaftlicher Leistungen ist ungenau, mit subjektiven Mängeln behaftet.

Dies wird z. B. daran deutlich, dass die gleiche Leistung zu verschiedenen Zeitpunkten höchst unterschiedlich bewertet wird - sogar von denselben Bewertern, Kritikern. Dazu eine ausführliche Darstellung von Peters und Ceci (1977):

Die Korrelation der Bewertungen durch Gutachter in sozialwissenschaftlichen Zeitschriften, vor allem psychologischen, bei dem *Journal of Personality and Social Psychology* erwiesen sich nur als gering bis mittelhoch, höchstens + 0,55. Für Gutachter von eingereichten Manuskripten betrug sie nur 0,26. Keine Interrater-Übereinstimmung wurden bei Manuskripten an das *Personality and Social Psychology Bulletin* gefunden. Bei der Analyse von 1000 Gutachten über 500 Manuskripten, die beim *Sociological Quarterly* eingereicht worden waren, ergab sich hinsichtlich der Einstufung auf einer 5 - Punkte - Skala:

1/3 völlige Übereinstimmung

1/3 ähnliche Übereinstimmung

1/3 keine Übereinstimmung der beiden Gutachter.

3.2 Größe der Universitäten

Gutachter von großen Universitäten bewerteten Artikel signifikant besser, die von großen Universitäten stimmten als von kleineren, weniger prestigereichen Universitäten. Gutachter von kleineren Universitäten zeigten diese unterschiedliche

Bewertung nicht. Artikel, die von Autoren aus kleineren Universitäten kamen wurden häufiger positiv von Gutachtern bewertet, die ebenfalls von kleineren Universitäten kamen, als von Gutachtern aus großen Universitäten. Im Vergleich zu Gutachten aus kleineren Universitäten bewerteten Gutachter aus größeren Universitäten Artikel positiv, die von Autoren großer Universitäten eingereicht wurden.

3.3 Bestätigungsneigung

Artikel, die die eigene Position vertraten, wurden positiver bewertet, Artikel negativer bewertet, die der eigenen theoretischen Position entgegenstanden.

3.4 Zeitlich unterschiedliche Bewertung des gleichen Manuskripts

Da Herausgeber oft darauf hinweisen, dass das wissenschaftliche Manuskript von zwei Gutachtern geprüft worden ist, die dieses Forschungsgebiet gut kennen, machten Peters und Ceci (1977) ein interessantes Experiment. Aus 11 weit verbreiteten Fachzeitschriften wählten sie 11 neuere Artikel aus, die von Autoren prestigereicher Universitäten stammten.

Peters und Ceci veränderten den Titel des Originalartikels, die Zusammenfassung und die Anfangssätze der Einleitung wurden verändert. Und an Stelle des Originalautors wurde ein Pseudonym aus einer fiktiven Institution gewählt. Inhaltlich wurde nichts verändert.

90 % der Herausgeber und Gutachter erkannten nicht, dass diese Pseudomanuskripte bereits erschienen waren. Nur 18 % empfahlen die Veröffentlichung in ihrer Zeitschrift. Die anderen lehnten es ab, wegen methodologischer, theoretischer und stilistischer Probleme - obwohl der Originalartikel hinsichtlich methodologischer, stilistischer u. a. Kriterien genau überprüft worden war!

Aus einer anderen Perspektive gesehen, zeigt diese Untersuchung, dass man befürchten muss, dass manche wissenschaftliche Erkenntnisse nur deshalb Schwierigkeiten bekommen, sich durchzusetzen, weil sie nicht von einer Person mit Prestige stammen.

4. Gibt es ein „Gesetz des wissenschaftlichen Wettbewerbs“?

Ich habe nur einige der subjektiven Faktoren geschildert, die die Bewertung eines Vortrages oder einer wissenschaftlichen Leistung beeinflussen (eine ausführliche Besprechung s. Füllgrabe, 1978). Das Hauptgewicht liegt hierbei auf dem „Dr. Fox – Effekt“, der zwei Seiten hat, eine positive und eine negative. Die positive Seite besteht – wie Ware und Williams in mehreren ihrer Artikel (1975, 1977) ausdrücklich betonen- darin, dass durch eine lebendige Darstellung das Interesse für ein Thema oder Sachgebiet geweckt werden kann.

Die negative Seite besteht aber darin, dass sogar informationsarme Darstellungen Begeisterung und den Eindruck von Kompetenz und Wissenschaftlichkeit auslösen können. Offensichtlich gibt es relativ viele Menschen, die kein „Messinstrument“ besitzen, um die Richtigkeit einer Aussage oder ihren Informationsgehalt abschätzen zu können, oder sie bemühen sich nicht, ein solches Messinstrument zu entwickeln, vielleicht, weil sie unüberprüft von der Meinung ausgehen, dass sie automatisch richtig urteilen könnten.

Es wäre aber sinnvoll, ein solches Messinstrument zu entwickeln, weil es viele Thesen gibt, die überzeugend klingen und erhebliche soziale und politische Auswirkungen haben, obwohl diese Thesen undifferenziert oder sogar falsch sind. Ein anschauliches Beispiel ist die These, dass „Intelligenz zu 80% angeboren“ sei und die damit verbundene, unausgesprochene Botschaft, dass es sich nicht lohne, Intelligenz zu fördern. Bezeichnenderweise war es ein bedeutender **Genetiker, T. Dobzhansky** (1973), der diese These aus genetischer Sicht scharf kritisierte und einige populäre Thesen zu genetischen Fragen folgendermaßen bewertete: „... um so irreführender, als es oberflächlich betrachtet so plausibel ist.“ Er weist z.B. (S. 27) darauf hin, dass es sich bei den imponierend klingenden Zahlen (z. B. 80%) nicht um mathematisch exakte Berechnungen handelt, sondern um *Schätzungen* (S. 27) und „...daß Heritabilität keine innere Eigenschaft eines Merkmals ist, sondern eine Eigenschaft der Population, in der es auftritt (Dobzhansky, 1973, S. 28).“ Außerdem werden durch die Erbanlagen keine feststehende Körpergröße, Intelligenz usw. vererbt, sondern nur eine **Reaktionsbreite**, d. h. die mögliche Unter- und Obergrenze des Merkmals (s. a. Füllgrabe 1997). Anschaulich haben dies auch die beiden Genetiker Vogel und Propping (1981, S. 340) geschildert. Vogel und Propping (1981) weisen in diesem Zusammenhang auch darauf hin, dass die Daten, die der bekannte und angesehene Engländer Sir Cyril Burt berichtete und die häufig zur Stützung der These von der angeborenen Intelligenz herangezogen wurden, schlichte Fälschungen darstellen. An einer Stelle ihrer Ausführungen (S. 129) ist der Bezug zum „Dr. Fox – Effekt“ deutlich (ohne dass sie diesen Effekt aber selbst erwähnen).

Neben den Informationen, die sie geben und die zeigen, das unsere Schicksal keineswegs „mitgeboren“ ist, selbst bei ungünstigen Erbanlagen nicht, ist ihr Buch von wissenschaftstheoretischem Interesse. Vogel und Propping (1981) zeigen nämlich auch den Einfluss von Vorurteilen, dem Zeitgeist und anderen subjektiven Einflussgrößen auf das wissenschaftliche Arbeiten auf. Und sie verbreiten Optimismus mit ihrer These vom „**Gesetz des Wettbewerbs**“ (S.129): Man sieht aber, dass der „Dr. Fox – Effekt“ dieses Gesetz des Wettbewerbs verzögert, verhindert oder außer Kraft setzt. Deshalb ist es notwendig, dass es Menschen gibt, die leicht „Mist entdecken“ können (Ware & Williams, 1975, 1977, Naftulin et al., 1973). Dies ist äußerst wichtig, angesichts der Äußerung von Dobzhansky (1973, S. 9 – 10): „Zyniker behaupten, die Menschen verwendeten die Wissenschaft nur dazu, ihren selbstsüchtigen Interessen und Voreingenommenheiten den Anstrich des Respektablen zu geben.“

Wie die Untersuchungen von Naftulin et al. (1973) zeigen, gibt es durchaus Menschen, die die Informationsarmut der Vorlesungen von Dr. Fox durchschauten. Es ist leider nicht bekannt, welche Strategien diese Personen benutzen, um nicht dem „Dr. Fox – Effekt“ zum Opfer zu fallen. Man kann aber davon ausgehen, dass sie die Aussagen von „Dr. Fox“ mit realitätsorientierten Kriterien verglichen, gleichgültig, ob dies bewusst oder nicht bewusst geschah. Es ist auf jeden Fall wichtig, sich bewusst zu machen, auf welche Kriterien man sein Urteil stützt. So selbstverständlich, wie dies sein sollte, die „Dr. Fox“ – Studien zeigen, dass es viel zu selten gemacht wird. Dies ist aber **die** entscheidende Strategie, um den Informationsgehalt einer Information abschätzen zu können.

5. Das Wissenschaftsspiel

Die Äußerungen von Vogel und Propping (1981) zeigen, dass es sinnvoll ist, Wissenschaft auch spieltheoretisch zu betrachten. Auch Agnew und Pyke (1969) formulierten auf dem Rückumschlag ihrer Buches eine spieltheoretische Betrachtungsweise: „Möchten Sie ein Spiel spielen? Es wird genannt „Woher weißt Du das?“. Sie spielen es jeden Tag. Wenn Wissenschaftler es spielen, nennt man es das **Wissenschaftsspiel**.

Auf S. VII schreiben sie: „... es ist klug für uns alle, mit diesem wichtigen Spiel vertraut zu werden – nicht nur, um seine Stärken zu schätzen, sondern auch, um seine menschlichen Grenzen zu verstehen. Denn wissenschaftliches Arbeiten ist wie andere Spiele eine Mischung aus Kunstfertigkeit, Unternehmungsgeist und Erfindungen, die lose durch von Menschen gemachte Regeln zusammengehalten werden. Die Bestandteile jeden großen Spiels können (auch) in der Wissenschaft gefunden werden: massive Anstrengungen, Ziele, Tore, große Spiele, mittelmäßige Spiele, lausige, schlechte Spiele, Schiedsrichter, die klare Entscheidungen treffen, einseitige, benachteiligende Entscheidungen und natürlich falsche Entscheidungen. Es gibt Preise und Strafen. Sie werden Integrität finden, Würde und Handel, zusammen mit Glück und Pech – vor allem finden Sie aber Engagement. Deshalb werden Sie, wenn Sie ein großes Spiel lieben, die Wissenschaft lieben, als Spieler oder als Zuschauer. Wenn Sie dagegen nur auf einem ebenen Feld spielen wollen, an einem hellen Tag, mit unabhängigen Schiedsrichtern, die genaue Regeln anwenden, dann ist Wissenschaft vermutlich nicht Ihr Spiel“ (Agnew & Pyke, 1969, pp. VII – VIII).

Wie irrational aber dieses Spiel sein kann, zeigen die erwähnten Analysen und Untersuchungen von Peters und Ceci (1977).

Literatur:

- Agnew, N. McK. & Pyke, S. W. (1969). *The science game*. Englewood Cliffs: Prentice Hall.
- Ceci, S. J. & Peters, D. P. (1982). Peer review: A study of reliability. *Change*. September 1982, pp. 44 - 48.
- Dobzhansky, T. (1973). *Intelligenz: Vererbung und Umwelt*. München: mgv.
- Füllgrabe, U. (1978). *Menschenkenntnis*. Stuttgart: Richard Boorberg Verlag.
- Hyman, R. (1977). „Cold reading“: How to convince strangers that you know all about them. *The Zetetic*. Spring/Summer 1977, pp.18-37.
- Naftulin, D. H., Ware, J. E. & Donnelly, F. A. (1973). The Doctor Fox lecture: A paradigm of educational seduction. *Journal of Medical Education*, Vol. 48, pp. 630 – 635.
- Peters, D. P. & Ceci, S. J. (ohne Jahr). *A naturalistic study of the journal review process in psychology*. Manuskript.
- Peters, D.P. & Ceci, S. J. (1981). Peer review practices of psychological journals: The fate of accepted, published again. *The Behavioral and Brain Sciences*, pp.1 – 25.
- Psychologie Heute* (1980). Reden Sie mal höheren Blödsinn. August 1980, S. 7 – 8.
- Schuler, H. (1972). *Das Bild vom Mitarbeiter*. München.
- Snyder, C.R. & Clair, M. (1976). Effects of expected and obtained grades on teacher evaluation and attribution of performance. *Journal of Educational Psychology*, Vol. 68, Nr. 1, pp. 75 – 82.
- Vogel, F. & Propping, P. (1981). *Ist unser Schicksal mitgeboren?* Berlin: Severin und Siedler.
- Ware, J. E. & Williams, R. G. (1975). The Doctor Fox effect: A study of lecturer effectiveness and ratings of instruction. *Journal of Medical Education*, Vol. 50, February 1975, pp. 149 – 156.
- Ware, J. E. & Williams, R.G. (1977). Discriminant analysis of student ratings as a mean for identifying lecturers who differ in enthusiasm or information – giving. *Educational and Psychological Measurement*, Vol. 37, Nr. 3, pp. 627 – 639.